DOCUMENT RESUME

ED 321 527 FL 018 597

AUTHOR Pico, Eliseo
TITLE Issues of Reliability and Validity of the Partial Dictation Test as a Measure of Listening Comprehension.
PUB DATE Apr 90
NOTE 17p.; Paper presented at the Annual Meeting of the World Congress of Applied Linguistics (9th, Casandra-Halkidiki, Greece, April 15-21, 1990). Photoreduced type, page 12, will not reproduce legibly.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cloze Procedure; *Dictation; *English (Second Language); Foreign Countries; High Schools; Language Research; *Language Tests; *Listening Comprehension; *Test Reliability; *Test Validity; Verbal Tests
IDENTIFIERS Spain

ABSTRACT

A study investigated the reliability and validity of the partial dictation test, a combination of dictation and cloze procedure, to measure foreign language listening comprehension. Subjects were 78 students of English as a Foreign Language in the second year of high school in Spain. Students listened to a passage from a commercial reader three times; no pauses were used in the reading. The test was administered in the classroom. The text was brief (268 words), with 30 words deleted. Reliability was tested using five methods, revealing high internal consistency. Measures of content and predictive and concurrent validity were also applied. Results suggest relatively short dictation tests can be constructed to provide useful measures of classroom second language proficiency. Among the practical advantages seen in this adaptation of dictation are ease and speed of scoring and adaptability to program needs and objectives. Further research is recommended on the validity of the procedure, and to determine whether partial dictation can discriminate among different terms in a course, or whether the desirable properties are shown by other dictations based on other text types, scoring and administration procedures, language modalities, segment length, and deletion decisions. A brief bibliography, a copy of the test, and tabulations of results are appended. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED321527

ISSUES OF RELIABILITY AND VALIDITY OF THE PARTIAL DICTATION TEST AS A MEASURE OF LISTENING COMPREHENSION

Eliseo Picó

Abstract

The partial dictation is a technique to measure foreign language proficiency in listening comprehension. Students listen to a passage recorded on tape while they write down the words missing in their answer sheets. This paper describes the test specifications and reports results from an administration to BUP2 students. The paper also reports the findings of a validity study which involved the use of two latent trait models, the S-P model (Harnish, 1983), and the Rasch model (Wright and Stone, 1979).

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E. Pico

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL018597

BEST COPY AVAILABLE

ISSUES OF RELIABILITY AND VALIDITY OF THE PARTIAL DICTATION TEST AS A MEASURE OF LISTENING COMPREHENSION

Eliseo Picó

Paper presented at the
9th World Congress of Applied Linguistics
Casandra-Halkidiki, 15-21 April 1990

ISSUES OF RELIABILITY AND VALIDITY OF THE PARTICAL DICTATION

Eliseo Picó*

Introduction.

Integrative theories of testing (Oller, 1979) have lent new support for partial dictation as a method of assessing learner comprehension. In this approach to listening comprehension, students are required to write several words at a time in blanks left in a text which they have just heard. This technique is, then, a combination of dictation and the cloze procedure. However, little research has been conducted on partial dictation as a measure of global listening comprehension or as a research tool.

On the other hand, Boyle (1987) has gathered factor analytic evidence to support the separate nature of the listening comprehension factor in language ability. If listening comprehension is a separate factor of second language proficiency, efforts have to be made to measure that factor accurately and efficiently. The following study was carried out to validate a listening comprehension test of foreign language proficiency.

Review of the literature.

Only three studies involving partial dictation and second language learners have been published to date, Johansson (1973), Chaudron (1985) and Pinsent (1986).

Valette (1967) was the first one to suggest a partial or spot dictation test. She mentioned two advantages of partial dictation tests: on the one hand, it is easy to administer and easy to score; on the other hand, it permits the teacher to test problem areas. The first of the mentioned advantages is really one of the strengths of partial dictation. But the second advantage may not be an advantage at all. We are not sure whether problem areas can be tested with partial dictations. But there is evidence to suggest (Picó 1989) that very simple words or expressions may turn out to be very difficult depending on their context of situation. Predictions made by teachers concerning the relative difficulty of the testing items in a partial dictation may encounter no empirical support from the students' responses.

While some researchers have used cloze procedures as suggested by Taylor (1953) following oral presentation of the text (Smith and Rafiqzad 1979, Oakeshott-Taylor 1979, Henning, Gary and Gary 1981) they all involved the deletion of just one word in every blank. We reserve the expression "listening cloze" for that type of test, while we will consider "partial dictation" that test involving deletions of more than one word which are filled while listening to the recorded text.

Following Valette (1967), the first operationalization of the partial dictation in the strict sense was offered by Stig Johansson (1973). He conducted the first published research on the use of partial dictation as a test of foreign language proficiency. This was the first detailed account of the partial dictation to be found in the professional literature. His format differs from ordinary dictation in that firstly, the passages presented for dictation are played from a tape recorder, that is not read out by the teacher, and secondly, subjects listen to the recording only once.

Partial dictations are also different from cloze. They overcome the local redundancy problems inherent in cloze tests pointed out by Carroll (1972). Using multiple word deletions, "the restorations of the missing elements would have to rely much more on the comprehension of the whole text" (Johansson 1973: 126). Partial dictations differ also from cloze in that subjects listen to the complete text. Johansson claims that the listening element should overcome Carroll's criticism that cloze measures general intelligence factors by testing inferential processes since it tests aural comprehension within a limited time span.

Johansson's operationalization of the partial dictation has been criticized (Klein-Braley, 1985) on the grounds of validity and his study has received severe criticism because of Johansson's failure to provide validity data. Johansson's use of statistics has been found puzzling (Klein-Braley, 1985) as he did not present the full correlation matrix. Mean scores and standard deviations were also missing. He left unexplained the inclusion of only 20, allegedly randomly selected, students in the correlations when 80 students took the test.

On the other hand, Pinsent (1986) found that the internal reliability for his partial dictations was high (ranging from .71 to .96 estimated with Cronbach's alpha), but they did not correlate with standardized tests of global comprehension. In other words, partial dictations have reliability but lack concurrent validity.

Chaudron (1985) indicated that the partial dictation is a reliable measure. Alpha coefficients ranged from .77 to .89 for his two experimental conditions. Besides, he indicated that the partial dictation technique can be interesting with the purpose of researching certain input objectives.

In a series of studies on the influence of the L1 in the acquisition of the L2 carried out in Finland, Hakan Ringbom (1987) has also used the partical dictation procedure. He indicated that the partial dictation procedure was sensitive to differences in L1. Kuder-Richardson formula 21 estimates calculated by the author of this paper on Ringbom's data ranged from .64 to .89.

Our operationalization.

The procedure we have used to design and administer the partial dictation differs somewhat from previous operationalizations. The most important difference is that the

students listen to the passage three times (emulating the dictation standard procedure): one for global understanding (without text), a second one to allow students to write down their answers, and a third one to allow students to check what they have written.

A second difference from Johansson's is that there are no pauses on the tape. In Johansson's operationalization, as students only listened to the passage once, pauses to allow students to write had to be inserted. In our operationalization, students have ample time to write down their responses, leaving the recording untouched.

A third difference from Johanssons's is that our test is carried out in the ordinary class (provided the room does not present acoustic problems), not in the language laboratory. There is no need for headsets. Though it can also be used with headsets.

A fourth difference is related to test length and text length. Johansson used longer texts than we did and he deleted one hundred words. We have used shorter texts having thirty deletions. Our partial dictation (See Appendix, Protocol) contained 268 words, the intention being that the test should be presented in just one face of a page. Thirty words were deleted forming eight chunks.

M E T H O D

Subjects.

Our sample consisted of 78 students of English as a foreign language taking mandatory classes in the second year of their secondary (high) school studies (the equivalent of American tenth graders; in Spain: Segundo de BUP). They were in two different classes (2A, n = 38; 2B, n = 40). They were all students of this author at IB Baldiri Guilera (El Prat de Llobregat, Barcelona) at the time of administration of the main test (March 1987). They also completed a battery of tests. Their ages ranged from 15 to 17. They were all Spanish-speaking children possessing differential degrees of bilingualism Catalan/Spanish. The amount of years they had been studying English ranged from two to seven. Their proficiency ranged from elementary to high intermediate.

Our subjects completed the test as part of the listening component of the battery to be used for their term grades.

Instrument.

The partial dictation we used for this validation study was obtained from passages taken from Five Ghost Stories (London: Heinemann) a commercial reader appropriate for the proficiency level of our students. A sample copy of the protocol is to be found in the Appendix. Thirty words were deleted forming eight segments. We used a recorded reader as starting point. We thereby skipped the problem of finding a native speaker to record the passage.

To score our test we adopted a simple criterion: one point per correct word restaured. Minor spelling mistakes (i.e. not affecting the meaning of a word) were considered as correct.

Administration.

It has already been said that Johansson (1973) recommends that the administration be carried out in a language laboratory; today's reproduction quality of certain cassette players makes this requirement unnecessary, provided that the room does not present problems of acoustics. A CROWN cassette player was used in our administration.

The students completed the test as part of the listening component of their regular testing battery for term evaluation. Testing time lasted less than fifteen minutes.

Reliability.

A set of classical and IRT analyses was performed to investigate the reliability of the partial dictation test. Reliability data are given in Table 1.

The internal consistency was determined by applying five formulas. The first one was the Kuder-Richardson formula 20, an index of the homogeneity of a set of binary-scored test items (i.e. items that can be scored as right or wrong). The figure yielded was .91, which is very high for a language test.

The second one, the Kuder-Richardson formula 21, which is an algebraic derivation of K-R20, assumes that the p values of all items (i.e. the proportion of examinees passing an item) are identical. If that is not case, then K-R21 yields a slightly lower estimate of reliability than K-R20. Our partial dictation yielded a K-R21 of .87, which is also quite high.

A generalization of K-R20, known as Cronbach's alpha, yielded a reliability of .91. It can be noticed that Cronbach's alpha approach to estimation of internal consistency yields the same results as the K-R20.

A fourth method of estimating reliability was the split-half. The test was divided in two parts, odd-even items, and the two subscores were correlated and then applied the Spearman-Brown formula. The reliability reached the value of .94, which indicated very high internal consistency.

The fifth method was the Guttman split-half reliability. This method does not assume equal variance for the two halves. In the case the two variances are equal the figure obtained is the same as the Spearman-Brown (Magnusson, 1967). This is our case. The figure yielded was .94. We know, then, that the variances for the two halves are equal.

In a manner of summary, it can be indicated now that the partial dictation procedure shows high internal consistency

reliabilities, even with a short test like the one we administered.

Validity.

Test validity is generally considered as the single most important aspect of a test. Validity can be defined as the degree to which a test measures what it is supposed to measure (Popham, 1981). The three types of validity that may be important in relation with the partial dictation may be content validity, predictive validity and concurrent validity.

Content Validity

Maybe the best indicator of the validity of a test is obtained by examining the content of the test. We have to ask ourselves if the items specify the domain we are trying to measure, in other words if out sample is a representative sample for the entire domain so that it can be legitimate to generalize from the performance on the sample of items in this particular test to performance over the entire domain. The adequacy of the sampling can be checked by examining the agreement between expert judges. A group of 22 expert teachers and researchers, native and non-native, massively found appropriate the present procedure (Picó, 1989).

The favourable opinion of experts provides certain empirical evidence of the validity of the partial dictation procedure.

Predictive validity

In evaluation situations. it is important that a test have predictive validity. Predictive validity is defined as the degree to which a test helps to predict future performance. For 33 students who took two partial dictations within a 3-month interval (see "Intercorrelation table") the Pearson r was .64 (significant beyond .01). The correlation between the partial dictation and a test of language proficiency administered in a 3-month interval was .66 (significant beyond .01).

We can conclude from the above that the partial dictation test seems to have quite high predictive validity at least over a three month interval.

Concurrent validity.

In most test manuals a number of concurrent validity studies are presented. This type of validation consists of "checking the test score against some other observation that serves as criterion" (Cronbach, 1971). Collecting such data presents a serious problem for English as a foreign language because very few valid measures of foreign language competence are available.

In our case we have used the SLEP Listening Section as a criterion (see "Intercorrelation table"). We can consider the figures for the partial dictations and the Listening section of the SLEP. We have correlations of .53 and .54. These figures have

not been corrected for attenuation (Henning, 1987). It we correct them we get a correlation of .67 for the partial dictation "The Ghost" and a correlation of .68 for the partial dictation "The Passenger", both moderately high and statistically significant beyond .01.

Cziko and Lin (1984) found that log ability scores provided by Rasch analysis correlated in general more highly with other measures of language proficiency than did the raw scores. This is also our case (See "Intercorrelation table"). We see that the Rasch Log Ability (Wright and Stone, 1979) correlates more highly than the raw scores. We have a correlation of .61 for the partial dictation "The Passenger" and a correlation of .54 (both unattenuated) for the listening comprehension section of the SLEP. The correlation with the Reading Comprehension Section of the SLEP is very high (r = .74). This result is surprising as it can be interpreted in the direction of the partial dictation tapping common variance with reading comprehension in a greater amount than listening comprehension, as measured by SLEP.

A further indicator of validity is the disparity coefficient obtained from S-P curve analysis (Harnish, 1983). The disparity coefficient indicates the extent to which the set of items is excessively heterogeneous or that the group of students consists of two or more subgroups having varying degrees of exposure to the material being tested. As a rule-of-thumb it is said (Harnish y Romy, 1985) that a disparity coefficient of around .5 is "about right", while a value exceeding .6 is a danger signal. In the partial dictation "The Ghost" the disparity coefficient is .3, well below the danger signal. We can say, then. that there is no disparity between the students' ability and the difficulty of the set of items. Our set of items is not heterogeneous and the group of students is not heterogeneous either. Further comments on item performance can be found in Picó (1988). Part of the output from the S-P analysis can be found in the Appendix.

Discussion and conclusions.

The results of this study provide additional evidence that relatively short dictation tests can be constructed to provide useful measures of classroom second language proficiency.

Among the practical advantages of this adaptation of the dictation procedure are the ease and speed with which these tests can be scored in comparison to traditional dictation scoring methods which require attention to each individual word in a test passage.

It can be concluded that the partial dictation procedure is a good measure of second language proficiency which, apart from having the desirable statistical properties mentioned above, can nevertheless be easily adapted to match the particular needs and objectives of language teaching programs by choosing appropriate target materials from scientific texts to narrative novel or colloquial dialogue.

However, further research is needed to continue to

investigate the validity of this procedure and to determine whether partial dictation can discriminate among different terms in a course, whether the desirable properties are shown by other partial dictations based on other types of texts, scoring and administration procedures, language modalities, segment length, and deletion decisions.

NOTE

* This study has been supported by the Institut de Ciencies de l'Educació of the Universitat Politècnica de Catalunya in Barcelona. I wish to thank Craig Chaudron for his helpful comments, suggestions, and criticism on an earlier draft. Responsability for the final version of the paper, however, rests with the author.

References.

Boyle, J.P. (1987) Is listening comprehension a separate factor in language ability? IRAL 25.3: 238-256.

Chaudron, C. (1985) A method for examining the input/intake distinction. In S Gass and C. Madden (Eds.) Input and Second Language Acquisition. Rowley, Massachusetts: Newbury House, pp.285-299.

Cziko, Gary A. & Lin, Nien-Hsuan Jennifer (1984) The construction and analysis of short scales of language proficiency: Classical psychometric, latent trait and nonparametric approaches. TESOL Quarterly 18.4:327-347

Cronbach, L.J. (1971) Test validation. In R.L. Thorndike (Ed.) Educational measurement. Washington, D.C.: American Council on Education.

Harnish, D.L. (1983) Item response patterns: Applications for educational practice. Journal of Educational Measurement 20.2: 191-206.

Harnish, D.L. & Romy, N. (1985) SPP: Student Problem Package on the IBM-PC. University of Illinois at Urbana-Champaign: Institute for Child Behavior and Development.

Henning, G. (1987) A guide to language testing: Development, evaluation, research. Rowley, MA: Newbury House.

Henning, G., Gary, N. & Gary, J.O. (1981) Listening recall: A listening comprehension test for low proficiency learners. Paper presented al AILA, Lund, August 9, 1981.

Johansson, S. (1973) Partial dictation as a test of foreign language proficiency. Swedish-English Contrastive Studies, Report number 3. Lund, Sweden: Department of English, Lund

University.

Klein-Braley, C. (1985) Reduced redundancy as an approach to language testing. In C. Klein-Braley & U. Raatz (Eds.) C-Test in des Praxis. Bochum: AKS.

Magnusson, D. (1967) Test theory. Reading, MA: Addison-Wesley.

Oakeshott-Taylor, J. (1979) Cloze procedure and foreign language skills. IRAL 17: 150-158.

Oller, J.W. (1979) Language Tests at School. London: Longman.

Picó, E. (1988) Análisis de respuesta de ítems: El model de D.L. Harnish aplicado a pruebas de lengua extranjera. Paper presented at Sexto Congreso AESLA, Santander 13-16 April.

Picó, E. (1989) Alumnos y profesores ante las pruebas: ¿Percepciones desiguales? Paper presented at VII Congreso AESLA, Sevilla 5-7 April.

Pinsent, P.J. (1986) A comparison of tests of recall as measures of foreign language listening comprehension. M.A. thesis, University of Hawaii at Manoa, ESL Department.

Popham, W.J. (1981) Modern Educational Measurement. Englewood-Cliffs, NJ: Prentice-Hall.

Ringbom, H. (1987) The role of the first language in foreign language learning. Clevedon: Multilingual Matters.

Smith, L.E. & Rafiqzad, K. (1979) English for cross-cultural communication: the questions of intellibility. TESOL Quarterly, 33.3: 371-380.

Taylor, W.L. (1953) Cloze procedure: a new tool for measuring readability. Journalism Quarterly, 30: 415-433.

Valette, R.M. (1967) Modern Language Testing. New York: Harcourt Brace.

Wright, B.D. & Stone, M.H. (1979) Best Test Design: Rasch Measurement. Chicago: MESA Press.

APPENDIX

Test protocol and key

PARTIAL DICTATION
"The Ghost"

Name: _____
Class: _____
Teacher: _____

Instructions

Write down the missing words as you listen. You will listen to the passage two more times.

==========

He killed me quite easily. He banged my head on the stones of the road. Bang! God, I was wrong. All my hate went out with that first bang. I was wrong to make all that trouble because I found him WITH ANOTHER WOMAN . And now he was doing this to me -bang! That was the second one, and with it EVERYTHING went out.

My thin young ghost must have shone a little in the moonlight. I saw him look up from the body in a fixed sort of way. That GAVE ME AN IDEA . I would stay with him. All my life I was afraid of ghosts. Now I was one myself. And so now I would make someone else afraid. HE WAS NEVER AFRAID . He said there weren't such things as ghosts. Oh, weren't there! I'd soon teach him.

John stood up, still looking in front of him. I COULD SEE HIM CLEARLY . Slowly, all my hate came back. I put my face close up against his. But he didn't seem to see it, he just looked IN FRONT OF HIM . Then he began to walk forward. He seemed to be going to walk through me. I was afraid. But why should I -a ghost- be afraid OF HIS REAL BODY ? But fear doesn't act as you would expect, ever. And I moved back before him, and them moved to one side and let him pass. He was ALMOST LOST IN THE DARK street before I began to follow him.

But I don't think he could have escaped from me. There was still something between us.

Table 1

Reliability for the Partial Dictation "The Ghost"

Internal Consistency	
Kuder-Richardson formula 20	.91
Kuder-Richardson formula 21	.87
Cronbach's alpha	.91
Spearman-Brown split-half	.94
Guttman split-half	.94

Table 2

INTERCORRELATION TABLE

Test	1	2	3	4	5	6	7
1 PD Ghost		.64	.59	.53	.75	.66	.98
2 PD Passenger			.54	.54	.57	.61	.61
3 Entry test (Cloze)				.63	.63	.74	.58
4 SLEP List. Comp.					.57	.85	.54
5 SLEP Read. Comp.						.84	.74
6 SLEP Total (Scaled)							.67
7 PD Ghost Log Ability							

All correlations statistically significant
p<.01, two tailed, n = 33

Alpha critical = .34 for .05 significance level
= .44 for .01 significance level

School: IB "Baldiri Guitera"
Instructor: Eliseo Pico
Date: 03-29-87
Number of students: 40
Number of problems: 30

Student Number	Test Score (Raw)	Test Score (%)	Modified Caution Ind/Sgn	Problem Number 04	09	02	16	01	11	08	03	09	04	13	21	07	18	15	25	27	22	17	24	12	23	05	28	14	06	10	01
Answer Key				T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
A.LLUCH	30	100.0	0.00	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S
AVELLNO	30	100.0	0.00	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S
JA.SANZ	28	93.3	0.18	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	F	.	F	S	.	.		
E.MATaO	28	93.3	0.46	B	.	.	.	.	.	.	.	F	.	.	.	.	.	.	.	.	.	.	.	.	F	.	.	.	S	.		
SONIA.S	27	90.0	0.27	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	F	.	F	.	.	F	S	.	.		
D.GOMEZ	25	83.3	0.18	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	F	.	F	F	F	.	.	S	.	.	.	F	P
F.BELDA	25	83.3	0.12	A	.	.	.	.	.	.	.	.	.	.	.	.	F	.	F	.	.	.	.	.	.	.	S	.	.	F	F	P
JSIEIRO	25	83.3	0.34	A	.	.	.	.	.	.	F	.	F	.	.	.	.	.	.	.	.	.	F	.	.	.	S	.	F	.	P	
MANUSAN	24	80.0	0.09	A	.	.	.	.	.	.	.	F	.	.	.	.	.	.	.	.	.	.	F	.	.	.	S	.	.	F	F	P
MERMOSL	23	76.7	0.14	A	.	.	.	.	.	.	.	F	.	.	.	.	.	.	F	.	.	.	.	F	.	F	S	.	.	.	F	F
ALVAREZ	21	70.0	0.04	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	F	F	.	F	.	S	F	.	.	F	.	F	F	F
GONZALZ	21	70.0	0.09	A	.	.	.	.	.	.	.	.	.	.	.	F	.	.	.	.	.	.	F	.	S	F	F	.	F	.	F	F
CHIMENO	21	70.0	0.26	A	.	.	.	.	.	.	F	.	F	.	.	.	.	.	.	F	.	.	.	F	S	F	F	.	F	.	.	F
SEGARRA	19	63.3	0.15	A	.	.	.	.	.	.	.	F	.	.	.	F	.	F	.	F	F	.	S	F	.	.	.	.	F	.	F	F
GEMO.LP	19	63.3	0.22	A	.	.	.	.	F	F	.	.	.	.	.	F	.	.	.	.	.	.	F	S	.	.	.	F	F	.	F	F
BARRIGA	19	63.3	0.04	A	.	.	.	.	.	.	.	F	.	.	.	.	.	.	.	.	.	.	S	F	.	F	F	F	F	F	F	F
ABIGAIL	19	63.3	0.18	A	.	.	.	.	F	.	.	.	.	.	.	F	.	.	F	.	F	.	F	S	F	F	.	.	.	.	F	F
PORTERO	18	60.0	0.07	A	.	.	.	.	.	.	.	.	.	.	.	F	.	F	.	F	.	S	F	.	.	.	F	F	F	F	F	F
RAMIREZ	18	60.0	0.08	A	.	.	.	.	.	.	.	.	.	F	.	F	.	.	.	.	.	F	S	.	F	.	F	F	F	F	F	F
VILLAMV	18	60.0	0.04	A	.	.	.	.	.	.	.	.	.	F	.	.	.	.	.	.	.	S	F	.	F	F	F	F	F	F	F	F
M.GOMEZ	18	60.0	0.16	A	.	.	.	.	.	.	.	.	.	F	.	F	F	.	F	F	.	S	F	.	.	F	.	F	.	F	F	F
JALONSO	17	56.7	0.08	A	.	.	.	.	.	.	.	.	.	F	.	.	.	F	FSF	F	.	.	.	F	F	F	F	F	F	F	F	F
FARENAS	16	53.3	0.11	A	.	.	.	.	.	.	.	F	.	.	.	F	F	.	FS	.	.	F	F	F	.	.	F	F	F	F	F	F
ILLESCS	15	50.0	0.07	A	.	.	.	.	.	.	.	F	.	.	.	F	F	.	S	.	.	F	.	F	F	F	F	F	F	F	F	F
RODRIGU	15	50.0	0.10	A	.	.	.	.	.	.	.	F	.	.	F	.	FSF	.	F	.	.	F	F	F	F	F	F	F	F	.	F	F
DANILO.	14	46.7	0.22	A	.	.	.	.	.	.	F	F	.	.	F	F	FS	.	.	.	F	F	F	F	F	F	.	.	F	.	F	F
CALDERA	14	46.7	0.10	A	.	.	.	.	.	.	.	.	.	F	F	.	.	SF	F	F	F	F	F	.	F	F	F	.	F	F	F	F
JC.GARC	13	43.3	0.09	C	.	.	.	.	.	.	.	.	.	F	.	.	FSF	F	F	F	F	F	F	F	F	.	F	F	.	F	F	F
SAN.CAS	12	40.0	0.08	C	.	.	.	.	.	.	.	F	.	.	.	FSF	F	F	F	.	F	F	F	F	F	F	F	F	F	F	F	F
FSANCHZ	12	40.0	0.07	C	.	.	.	.	.	.	.	F	.	F	.	S	F	F	F	F	F	F	F	F	F	.	F	F	F	F	F	F
JMOLINA	12	40.0	0.04	C	.	.	.	.	.	.	.	.	.	.	.	FSF	F	F	F	F	F	F	F	F	F	.	F	F	F	F	F	F
A.ATUSO	10	33.3	0.05	C	.	.	.	.	.	.	F	.	.	FSF	.	.	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
GONZALZ	10	33.3	0.13	C	.	.	F	.	.	F	.	.	.	FS	.	F	.	F	F	F	.	F	F	F	F	F	F	F	F	F	F	F
HERREJN	9	30.0	0.21	C	.	.	.	F	.	.	.	F	FSF	F	F	.	F	F	F	F	F	F	F	F	F	F	.	F	.	F	F
VILARDB	9	30.0	0.18	C	.	.	.	F	.	.	.	F	S	F	F	F	F	F	F	F	F	F	F	.	F	F	F	F	F	F	F	F
J.MUBOZ	9	30.0	0.04	C	.	.	.	.	.	.	.	.	F	FSF	F	F	.	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
AM.VERA	8	26.7	0.30	C	.	.	.	.	F	F	.	FS	.	F	F	F	F	F	F	F	F	F	F	.	.	F	F	F	F	F	F	F
VIVANCS	8	26.7	0.01	C	.	.	.	.	.	.	.	FS	.	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
MACHORD	8	26.7	0.14	C	.	.	.	F	.	.	F	FSF	F	F	F	.	F	F	F	.	F	F	F	F	F	F	F	F	F	F	F	F
VALLEJO	5	16.7	0.09	C	.	.	.	F	FSF	F	.	.	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Problem Number				04	09	02	16	01	11	08	03	09	04	13	21	07	18	15	25	27	22	17	24	12	23	05	28	14	06	10	01

**** PROBLEM STATISTICS ****

Permuted Problem Number	Answer Key	Problem Total	Percent Correct	Modified Caution Ind/Sgn	
4	T	40	100.0	0.00	Y
9	T	40	100.0	0.00	Y
12	T	39	97.5	0.20	Y
6	T	36	90.0	0.02	Y
10	T	36	90.0	0.25	Y
11	T	36	90.0	0.15	Y
2	T	34	85.0	0.30	Y
8	T	30	75.0	0.11	Y
3	T	29	72.5	0.52	Z
19	T	28	70.0	0.05	Y
24	T	28	70.0	0.09	Y
23	T	26	65.0	0.20	Y
13	T	25	62.5	0.22	Y
7	T	23	57.5	0.02	Y
18	T	21	52.5	0.08	Y
21	T	21	52.5	0.05	Y
25	T	21	52.5	0.23	Y
20	T	18	45.0	0.06	W
22	T	18	45.0	0.04	W
27	T	17	42.5	0.20	W
1	T	16	40.0	0.20	W
26	T	16	40.0	0.18	W
29	T	16	40.0	0.11	W
30	T	16	40.0	0.13	W
15	T	15	37.5	0.12	W
28	T	15	37.5	0.11	W
14	T	10	25.0	0.18	W
16	T	9	22.5	0.04	W
5	T	7	17.5	0.03	W
17	T	6	15.0	0.00	W

**** STUDENT SUMMARY ****

Average raw score	= 17.30
Standard deviation of raw score	= 6.82
Average percent of items correct	= 57.67
Average modified caution index	= 0.13
Standard deviation of modified caution index	= 0.10

**** PROBLEM SUMMARY ****

Average item difficulty (%)	= 57.67
Standard deviation of item difficulty	= 10.07
Average modified caution index	= 0.13
Standard deviation of modified caution index	= 0.11

**** TEST SUMMARY ****

Average overall student performance on test (%)	= 57.67
Reliability coefficient (Cronbach's alpha)	= 0.91
Disparity coefficient	= 0.30